Bull. Fac. Agr., Saga Univ. No. 90 : 31～44 (2005)

Predicting Functionally Important Residues by Site Rate Shift Analysis and 3D Structural Mapping

Feng WANG[1], Hirohisa KISHINO[1,2] and Yasuhiko WADA
(Laboratory of Animal Production)
Received October 27, 2004

Summary

The function divergence in the ligand-binding domain of estrogen receptor (ER) alpha and beta paralogous proteins was analyzed by site rate shift method. The functional residues responsible for functional changes in ER subtypes were predicted by the posterior probability and likelihood ratio tests (LRTs). Combined with 3D structure mapping, we could further discriminate the function of the predicted residues, i.e. ligand binding or dimerformation interface. Therefore, this integrated approach could provide the new insights into the sequence-function relationship of ER family and is useful for screening the potentially functional residues as the experiment targets..

Key words: estrogen receptor, function divergence, function residues, structural mapping

Introduction

The estrogen receptor (ER) is a ligand-activated transcription factor that mediates the physiological effects of the female sex steroid hormone 17 beta estradiol (E2), and regulates the expression of genes involved in the growth, development and function of a diverge range of tissues.

The ER is a member of nuclear receptor (NR) super-family, which shares a common structural organization including six independent but interacting functional domains. The N-terminal A/B domain, C-terminal F domain and the D domain linking C and E domains are poorly conserved, thus provide limited phylogenetic information. The C or DNA binding domain (DBD) contains a two zinc finger structure, which is involved in binding of receptors to specific DNA sequences and in receptor dimerization. The DBDs of ER alpha and ER beta are highly homologous (>95%), thus, they can be expected to bind to various estorgen response elements (EREs) with similar specificity and affinity. The E or ligand-binding domain (LBD) is functionally complex since it mediates ligand binding, receptor dimerization, lignad-dependent transcription activation function and interactions with transcriptional coactivators and corepressors. The LBDs of ER alpha and ER beta share a modest degree of homology in their primary amino acid sequence (60%)[1,2]. Once bound by estrogens, the ER undergoes a conformational change, after which the receptor dimerizes and binds to ERE located in the promoters of target genes, and cause the

[1]BIRD, Japan Science and Technology Agency
[2]Laboratory of Biometrics, Graduate School of Agricultural and Life Sciences, University of Tokyo

changes in the transcription rate of these genes. However, the precise mechanism by which ER affects gene transcription is still poorly understood[3].

Two ER subtypes, ER alpha and ER beta, have been found in many vertebrate species, from teleosts to mammals[4]. In a previous phylogenetic analysis, the origin of ER gene duplication has been traced to a single duplication event at least 450 million years ago[5]. The parallel evolution of the two ER isoforms after this ancient duplication suggests that the two ER subtypes play unique roles in vertebrate physiology and reproduction, although they share a substantial degree of sequence identity, especially in DBD and LBD domain. The third subtype, ER gamma or ER beta2, has only been identified in teleosts from the recent works[4,6]. The more sampling in this group and the more information about the ligand-binding and structure features are needed to clarify the impact of the second duplication in the ER gene evolution[6].

Although related, ER alpha and beta are separate genes and code for proteins of differing lengths[7]. The distinct but often overlapping tissue distribution and ligand preference between two subtypes has been widely characterized[8,9]. Genistein and several other phytoestrogens have a significantly higher binding affinity for ER beta compared to ER alpha. Whereas endogenous estrogens and many other compounds have similar binding affinities for both subtypes. The estrogen-induced transcription activity of ER beta was less than that of ER alpha in some cells, the opposite was also found in other cells. The recent study indicated that ER beta inhibits ER alpha-mediated gene transcription in the presence of ER alpha, whereas, in the absence of ER alpha, it can partially replace ER alpha[10]. An important physiological role of ER beta is to modulate ER alpha-mediated gene transcription, supporting a "Yin Yang" relationship between ER alpha and ER beta, i.e. opposite but interdependent.

We are interested in the long-term evolution constraints operating on the function domains of ER subtypes after ancient duplication events. What kind of underlying mechanism shapes the different paralogous proteins with high degree homology but evident functional divergence, especially in LBD domain. Some mutagenesis experiments have identified the functionally crucial residues in the LBD, whose substitution between two subtypes or different species are related to the binding affinity difference. The human estrogen receptors ER alpha and ER beta, share only 56% amino acid sequence of the ligand binding domain (LBD), but the residues that surround the ligand are nearly identical[11].In fact, a single ER alpha point mutation (L384M) was largely sufficient to switch the diarylpropionitrile (DPN) response of this ER to that of the ERbeta type, but residues in helix 3 are also important in achieving the full ER beta selectivity to DPN[11]. The conservative changes between ligand binding pockets of human and rainbow trout ER alpha effectively exchanged phenotypic behavior of binding affinity, however, the unknown factors outside of the ligand-binding pocket are also involved for the complete exchange of the phenotypes[12].

To identify the patterns of conservation and variation at the individual position can help much for our understanding of selective constraints on the functional sites of the molecular[13]. The substitution rates vary from site to site within a protein and reflect the degree of different constraint. Proteins can evolve by selection of one or a few sequence changes that confer altered or novel function[14].

The sequence differences of ER subtype in LBD provide the molecular basis for subtype physiological function divergences. Thus, determining the functionally related residue substitu-

tions is very useful for developing subtype-selective ligands and pharmaceutical agents. Specially, it is important to explore the potential residues functionally related of the outside the ligand-binding pocket, as suggested by the known mutagenesis experiment[11, 12]. For identifying potential functional sites, it is better to employ the computational approach, because the intensive experimental mutation analyses are nearly impossible for many target sites.

The evolutionary rate changes at the sites of a protein implicated the possible changes in the selection and the changes in protein function. A significant rate difference between two subfamilies at a given site would thereby mean that there are different functional constraints in two groups. Therefore, it may indicate that the site is crucial for the functional changes in the two groups developed by the models to detect whether two subfamilies have undergone a functional divergence based on the rate changes at one site after gene duplication (type I function residues)[15, 16]. Knudsen *et al.* (2001, 2003)[17, 13] used the likelihood ratio tests (LRTs) to determine the significance of the rate differences at a specific site in different subfamilies. By the method, both the type I and type II function residues can be detected. Type I sites represent the amino acid conservation in one subfamily, but highly variable in another, implying that these residues have experienced altered functional constraints. Type II sites represent both amino acid conservation in two subfamilies, but with different biochemical properties, implying that these residues may be responsible for functional specification in two subfamilies. The evolution analyses, combined with structural information, has provided an effective way to predict function residues relating functional divergence between subfamilies[18]. In this study, we implemented this strategy to predict functional residues determining ER subtype specificity.

Materials and Methods

The ER, estrogen related receptor (ERR) and other steroid receptors (AR, PR, MR, and GR) amino acid sequences were collected from translated Genbank data bank by the keyword and BLAST search[19]. The possible Fugu homologous sequences were obtained by searching the predicted proteins of the fugu genome (http://www.ncbi.nlm.nih.gov/BLAST/Genome/fugu.html). The subtypes of fugu homologous sequences were designated by its position on the phylogenetic tree. The whole length of protein sequences was aligned by CLUSTAL X[20]. The ambiguous positions in the LBD were manually adjusted based on the structural information. The phylogenies of whole sequences and LBD sequences were respectively reconstructed with Neighbor-Joining, Maximum Parsimony and Maximum Likelihood methods using MEGA2[21] and TREE-Puzzle software[22]. The NJ tree topology of LBD amino acid sequences was largely consistent with that of whole length. This phylogeny was used for rate shift analysis on the ER subtype alpha and beta. The function residues were predicted by DIVERGE software[23] and the online sever made by Knudsen *et al.* (2003)[13]. The structure mapping and analysis of these predicted residues were conducted by VMD program[24].

The all the protein sequences used in this study are as follow: ER alpha for human (NP_ 000116), for mouse (P19785), for rat (P06211), for golden hamster (AAD53956), for pig (Q 29040), for horse (AAD17316), for sheep (AAK52104), for whiptail lizard (AAB35739), for green anole (AAC64412), for chicken (CAA27433), for quail (AAN63674), for zebra finch

(AAB81108), for Xenopus (P81559), for Atlantic salmon (AAG16713), for gilthead seabream (AAD31032), for blue tilapia (P50240), for nile tilapia (AAD00245), for bastard halibut (BAB 85622), for medaka (P50241), for Atlantic salmon (CAA61999), for rainbow trout (P16058), for fugu (SINFRUT00000062437), for goldfish (AAL12298), for zebrafish (AAK16740), for channel catfish (AAG24543), for North african_catfish (CAC37560), and ER beta for human (AAC 05985), for rhesus monkey (AF119229), for stump tailed macaque (AAK71317), for white tufted ear marmoset (CAA70546), for sheep (AAD55772), for cow (AAD24432), for pig (AAD45381), for rat (CAA05631), for mouse (AAB51132), for chicken (BAA88667), for starling (AAD56593), for quail (AAC36463), for seabream (AAD31033), for halibut (BAB85623), for Atlantic croaker (AAG16712), for Nile tilapia (AAD00246), for fugu (SINFRUP00000071164), for medaka (BAB79705), for goldfish (AAD26921), for carp (BAB91218), for rainbow trout (CAC06714), for zebrafish (CAC93849), for Japanese eel (BAA19851), for Atlantic croaker (AAG16711), for fugu (SINFRUP00000067205), for goldfish (AAF35170), for zebrafish (CAC93848), for channel catfish (AAF63157) and human AR (AAA51729), human GR (P04150), human PR (AAA60081), human MR (AAA59571), mouse ERR gamma (NP_036065), human ERR gamma (O75454), human ERR beta (O95718), mouse ERR beta (NP_036064), mouse ERR alpha (O08580), and human ERR alpha (CAA35778).

Results

1. The phylogenetic inference of ER based on amino acid sequences

The reconstruction of ER phylogeny based on whole length sequences was conducted using distance, parsimony and likelihood methods. For the distance-based tree, the orthologous sequences of ER alpha and beta were largely consistent with the vertebrate taxonomic relationships, and the paralogous proteins (alpha vs beta) resulting from gene duplication can also be recovered (Fig 1). The main branches of ER family were clustered as ((Teleosts alpha, Tetrapods alpha) (Teleosts beta, Tetrapods beta)). However, the MP and ML tree topology were like (Teleosts alpha, (Tetrapods alpha, (Teleosts beta, Tetrapods beta))). Obviously, this topology could not recover the species relationship in the ER alpha orthologous sequences. It is difficult to understand why tetrapods alpha branch is more closely related to ER beta than to Teleosts ER alpha. To accept the latter hypothetical tree, the gene conversion (recombination) assumption between alpha and beta subtypes is needed to interpret this evolutionary relationship. The null hypothesis for the evolutionary relationships of the multigene family members is that each orthologous gene member should reflect the evolutionary relationships of its species. Violation of this null hypothesis might indicate the gene conversion occurred between some members of a multigene family[25]. Here, ER alpha and beta could be seen as two gene members of ER family. Also, the inconsistent tree topologies between NJ and MP or ML inferred from the corresponding DNA alignment also exists. The phylogenetic trees by ML, distance method, and MP all supported monophy for each of Teleosts alpha, Tetrapods alpha, Teleosts beta, and Tetrapods beta. However, ER alpha (Teleosts alpha, Tetrapods alpha) did not make a monophyletic group in ML and MP trees. Distance method suggested monophyly of ER alpha, although the bootstrap support of 59% was not strong enough. For the empirical studies, it is important to use biological information in the final judg-

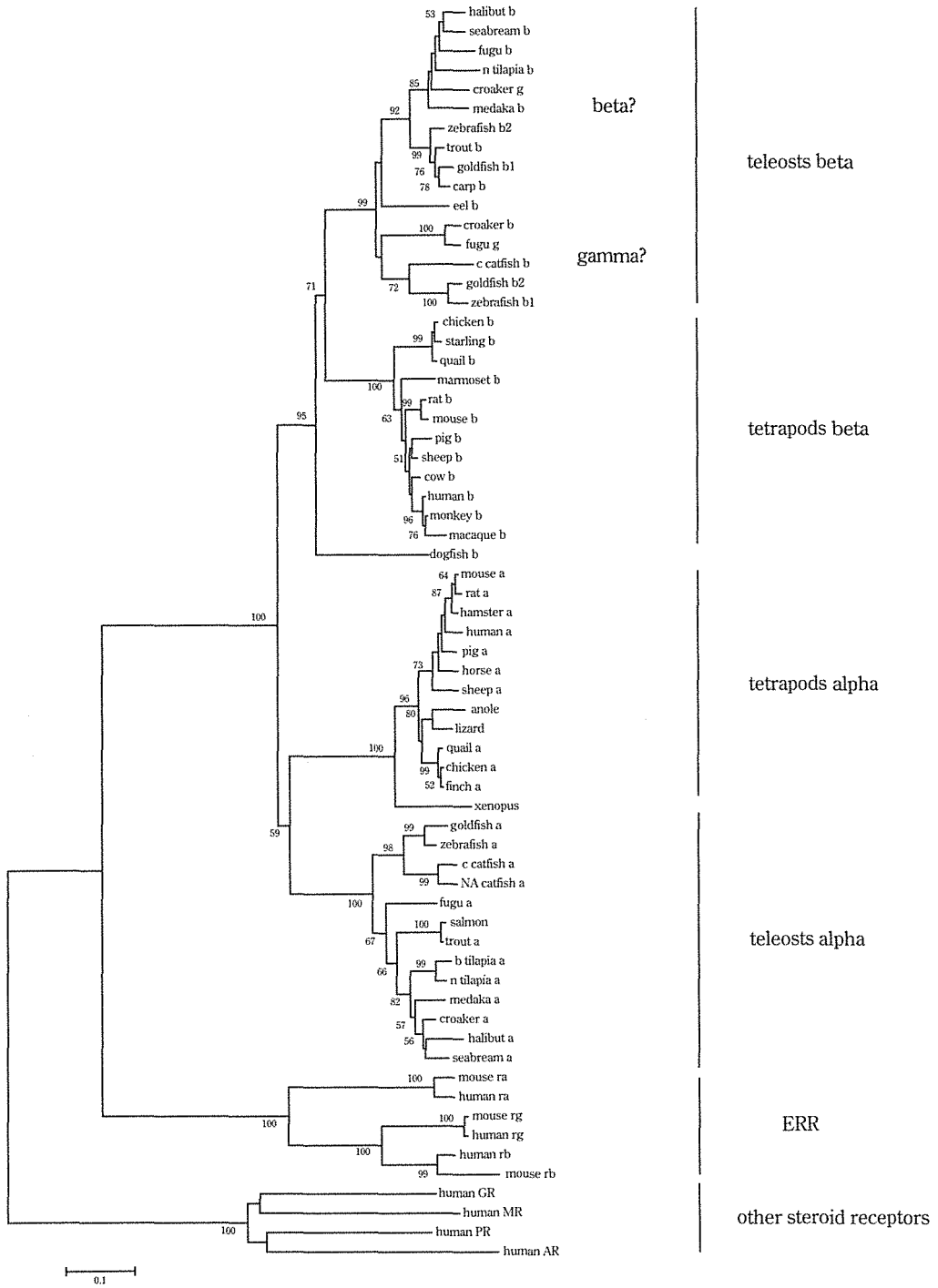

Fig. 1 The NJ tree of ER protein family and outgroup based on the whole length amino acids sequences. The names of the members are outlined. The numbers indicate the statistical support of the branch by the 1000 bootstrap replications (> 50% are shown). The vertical bar defined the outgroup (ERR and other steroid receptors) and sub-branches within the ER family. The possible two clades resulting from gene duplication in teleost lineage are labeled by ? symbol.

ment of the reliability of inferred trees[26]. In the previous studies[4,5], ER alpha and beta sequences are always kept in their own clades. Therefore, we assumed monophyly of ER alpha and used the distance-based tree topology (Fig. 1) for the analysis of functional divergence.

The NJ tree based on the Poisson Correction protein distance was shown in Fig. 1, with the gap handling option of complete delete in MEGA2. The internal branch support was calculated by 1000 bootstrap replicates. For the outgroups, the ERR branch was more close to ER branch than other steroid receptors (AR, GR, MR, PR). The ER family has the 100% statistic support for its monophyly. Within the family, the two branches, alpha and beta, have different robustness. The monophyly of ER alpha was less strongly support (59% vs 95% for ER beta). As mentioned above, the monophyly of ER alpha collapsed in the MP or ML inferred topologies. So, more works are needed to focus on the monophyly of the alpha clade in the evolutionary study of ER family. There are two main sub-branches in alpha and beta subfamily, respectively, teleosts vs tetrapods, all with the robust bootstrap support. Specifically, the dogfish beta was resolved in a more basal position in beta subfamily. This is reasonable because it belongs to the elasmobranch, the lower vertebrate. Because there is no corresponding sample in alpha subfamily, we deleted the dogfish data in the next analysis in order to keep the balanced sampling between two subfamilies. Within the ER beta subfamily, the possible gene duplication events in teleosts could be observed. The two paralogous sequences of goldfish, zebrafish, atlantic croaker and fugu are kept in two independent branches. Although the support for them is very low (less than 50%), this branching pattern is enough to support that the gene duplication events occur in teleosts lineage. Indeed, the third ER subtypes have been characterized by tissue distribution and ligand binding analysis[4]. It should be an open question awaiting clarification. In the present report, we are more interested in the functional divergence analysis after the first gene duplication in ER family, and because of the uncertain nature of ER gamma branch resulting from limited sampling, we only defined ER alpha and beta (including gamma) subtypes in the family.

2. Functional divergence analysis between the ER alpha and beta subfamily

The alignment of DBD domains and LBD domains were unambiguously compared with other domains of ER proteins. Considering the high percentage of sequence similarity (>95%) in DBDs, we only conduct the functional divergence analysis for LBDs of ER proteins. The phylogeny inferred only by LBD sequences was consistent with the whole length sequences. So, the alignment of ER LBD sequences was used to detect type I function changes by DIVERGE[23]. The coefficient of type I functional divergence θ between ER alpha and beta subfamily is 0.254400 ± 0.057323. It is significantly greater than 0, implying that altered functional constraints may take place at some amino acid residues after gene duplication. The posterior probability analysis was then conducted to predict important amino acid residues responsible for altered functional constraints (site-specific rate difference) between the two subfamilies. There are sixteen amino acid residues surpassing the cut-off value of posterior probability 0.5 (Table 1). Among the predicted residues, L384 in human alpha was predicted as type I residue (0.54) responsible for different functional specificity between alpha and beta subfamily. It was completely conserved in alpha orthologs and substituted to M in human beta. This site lied in the ligand-binding pocket. Sun *et al.* (2003)[11] have already recognized its roles for the ligand binding preference for human ER beta.

Table 1. Functionally important residues predicted by site specific rate shift

Human alpha/beta Type I and II residues defined by LRTs method		Type I residues defined by DIVERGE (posterior pr.>0.5)	Distances from ligand (or coactivator) of these residues 3ERD/1QKM (Å)
305 S/258 L	I		30.5/NA
313 D/266 E	I	0.72	27.5/27.0
319 L/272 L		0.58	18.6/8.2
327 L/280 V	I		9.1/9.0
328 Y/NA I			10.5/NA
329 S/281 L	II		9.8/10.4
330 E/282 I	I+II	0.79	12.8/9.8
338 S/290 T	I	0.87	13.0(5.6)/NA
344 G/296 M	II		8.2/8.5
345 L/297 S	II		8.8/8.4
356 H/308 H	I	0.59	11.2/11.2
373 H/325 F	I	0.99	23.3/25.0
382 A/334 C	II		12.2/11.9
384 L/336 M	I	0.54	6.4/6.3
389 I/341 M	I		7.9/7.7
398 H/350 H	I	0.77	15.1/15.7
432 S/384 T	I	0.77	12.6/12.6
437 M/389 E	I+II	0.81	21.5/21.1
441 Q/393 Q	I	0.63	19.1/19.5
442 G/394 H	I		16.9/17.6
445 F/397 Y	II		15.0/15.1
460 T/412 P	I		22.1/25.0
464 S/415 T	I	0.53	24.4/NA
468 S/419 D	II		25.7/NA
471 E/422 S	I+II		28.8/29.3
472 K/423 S	I		28.0/29.0
474 H/425 K	I+II		28.8/28.6
475 I/426 L	II		25.3/25.5
487 I/438 V	I	0.71	27.7/27.0
488 H/439 W	II		28.8/28.0
489 L/440 V	I		26.3/25.5
502 Q/453 M	I+II	0.71	29.9/28.4
509 L/460 M	II		20.1/21.5
517 M/468 A	I	0.70	8.7/8.0
522 M/473 M		0.55	6.2/6.3
526 Y/477 L	II		7.0/6.8
548 R/499 V	II		13.6(9.4)/24.3

In the first column, 391L, 522M are not predicted by LRTs method. In the second column, only the probability > 0.5 are shown. In the third column, the distances were measured between the alpha carbon of the predicted residues and the ligands or coactivators by VMD software.

The functional relevance of the other predicted residues will be evaluated in the structural mapping section.

3. LRTs methods for predicting type I and II functional residues

DIVERGE was mainly used to define the type I functional residues between two subfamilies. Knudsen *et al.* (2003)[13] employed a series of likelihood ratio tests to determine the type I and type

II functional residues. The type II functional residues were conserved in two subfamilies, but with different physical/chemical characteristics. This also indicates the different selective constraints in two subfamilies after gene duplication.

Thirty-five residues were predicted by the LRTs methods with distinct functional divergence between ER alpha and beta. Some of them were designated as type II functional residues. Others were type I or mixed of functional residues of type I and II. The type I residues defined by LRTs method were largely consistent with the prediction of DIVERGE analysis (Table 1).

4. The structure mapping of predicted residues by DIVERGE and LRTs

The predicted residues were all indicated on the 3D structure of human ER alpha and beta, the Protein Data Bank files are 3ERD and 1QKM, respectively. The LBD domain of human ER alpha was bound to synthetic nonsteroidial estrogen Diethylstilbestrol (DES) and the transcription coactivator GRIP1/CBM. The DES functions as agonist and promotes the coactivator recruitment[27]. The ER beta was bound to Genistein in a partial agonist manner[28]. The functionally divergent residues are randomly distributed on the 3D structure (Fig. 2). About 23 residues are in close contact with the bound ligands, which formed ligand-binding pocket. Between human ER alpha

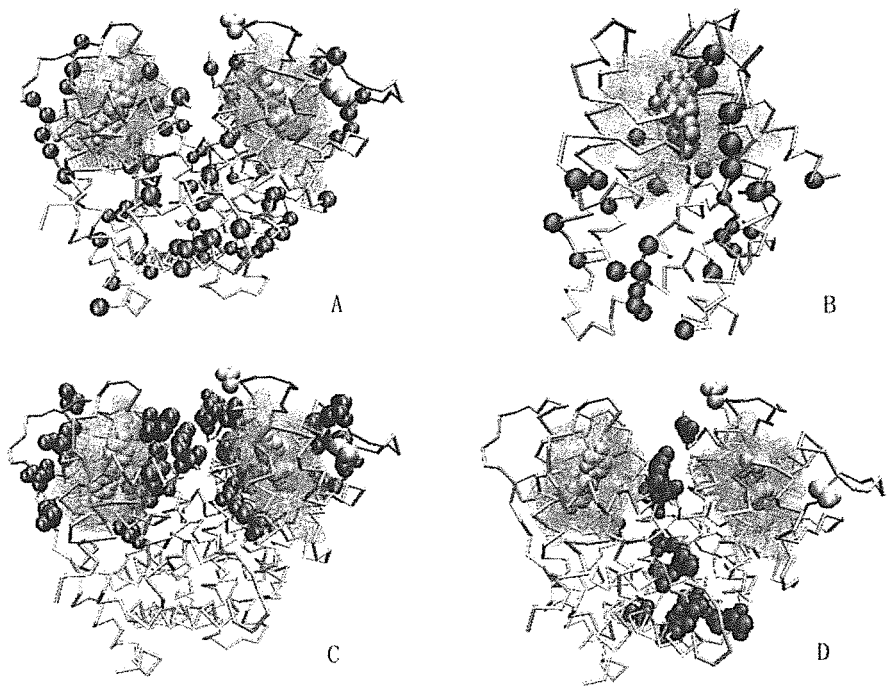

Fig. 2. The predicted residues were mapped on the ER LBD 3D structures. The ligands and coactivators are shown in Van Der Waals (VDW) form. The side chain residues at ligand binding pocket are dotted in grey color, enclosing the ligands. The C_α atoms of predicted residues in human ER alpha (3ERD, panel A) and human ER beta (1QKM, panel B) are shown in black color. The residues (black) with distances to ligands or coactivators within 10Å are presented in VDW form (panel C, 3ERD). The residues (black, VDW form) with significant differences of the solvent accessibility between the dimeric and monomeric form were distributed in the dimer interface (panel D, 3ERD). Original color figure of Fig. 2 is available on the World-Wide Web site of Prof. Y. Wada's laboratory in Saga Univeristy (http://genome.ag.saga-u.ac.jp/research/wang/fig2.jpg).

and beta, only two of them are different. L384 and M421 in ER alpha and M336 and I373 in ER beta. The structure analysis implicated that the two residues are major factors to determine ligand selectivity of ER alpha and beta[28]. Our evolutionary analysis failed to predict M421 I as functionally divergent residue. Besides, G344 and L345 (M296 and S297 in human beta) are predicted as type II functional residues by LRT methods (Table 1), which are away from the ligand compared with the residues at ligand binding pocket. It is possible for them to perturb the overall ER structure in a way that affects the interaction between the ligand and the residues nearby ligand binding pocket, T347 (299). Therefore, they may contribute to the ligand-selectivity of ER beta with an indirect contact to the ligand[11].

There are other mutant residues with altered transcription characteristics recorded in the mutation database of NucleaRDB (http://receptors.ucsf.edu/NR/). G442E mutation in human ER alpha have such effect[29], although it is not close to the ligand binding pocket. H373A mutation lacks the zinc-induced hormone release[30]. These mutation data do not support the functional alteration between ER alpha and beta, they are indeed indispensable for functional specificity of ER alpha (type I residues).

Enlightened by Sun *et al'* s (2003)[11] finding, the residues with indirect interaction to ligand can also contribute to ligand selectivity. We postulate such interaction have to depend on the nearby residue which is direct contact with ligand. So, the predicted residues near ligand or ligand binding pocket, would involve such indirect interaction, and are possible to affect ligand selectivity. Therefore, it is rational to set a possible interaction distance for screening such residues. One of the interaction distances were set as 10 Å used by Landgraf *et al.* (2001)[31].

Here, we choose 10 Å as interaction distance that was calculated between C_α atoms of the predicted residues and ligand or coactivator. By distance measure (Table 1), we can screen the residues with the distance less than 10 Å and map them onto the 3D structure (Fig. 2). These residues are more possible to be involved in the ligand−binding preference between ER alpha and beta. Maeda (2001)[32] calculated the differences between the solvent accessibility (SA) of a monomeric and a dimeric from of human alpha LBD structure. Among our predicted functional residues, seven of them have the differences larger than 0.1, i.e. 437M, 460T, 472K, 487I, 502Q, 509 L, 548R. They are possible to be involved in the dimerization. Mapping these residues were distributed on the interface between two monomers (Fig. 2). By the structure mapping, we discriminate the different involvement in functional roles of the residues predicted by the evolutionary analysis.

Disscussion

There are no evident clustering or special locations on the 3D structure for our predicted functionally divergent residues. This is different from the widely used Evolutionary Trace method, where the structured cluster in the protein structure is the criterion to identify the functional residue[33]. We used the evolutionary-based method to identify the function shift residues between subfamily. The statistic basis of this method makes it distinct from the simple list of conservative or variable residues in the alignment[16]. However, the dispersed substitution residues on the structure

suggested that any position in a protein could be important for the overall function[34]. The structural nature of functional residues makes it difficult to discriminate them from other non-functional ones.

The possibility of neutral mutations on the amino acid level indeed exists, evenly with distinct rate shift between paralogous proteins. Because the structure folds are more conserved than sequence level, therefore, at some positions, the residual substitutions can be tolerated without causing functional change[35]. It is not easy to distinguish the adaptive mutation (function-related substitution) from the neutral one. In the case of ER proteins, the predicted functional residues can be evaluated by the involvement in the ligand binding and dimer formation. By the distance measurement, we can discriminate the functional residues related to ligand binding. Those residues near the ligand are putative to be involved in the ligand binding specificity. And those residues distant to ligand, but distributed on the interface of dimer are more possible to be involved in the dimerization.

The most important functional domains of ER are DBD and LBD, which is indispensable for the ligand-activated transcription activation. At the nuclear receptor superfamily level, it is suitable to detect functional divergence in DBD between different receptor family[36]. We mainly focused on the functional divergence analysis of LBD of the subtypes derived by gene duplication in the evolution process of ER family. The enough calculations and structural information indicated that the evident functional divergence occurred between alpha and beta paralogs. The coefficient θ is significantly greater than 0, implying that altered functional constraints between alpha and beta may take place at some amino acid residues. The posterior probability and LRTs methods were employed to predict those residues responsible for the functional divergence. The predicted residues were further mapped on to the 3D structures to explore the functional residue distribution. Theoretically, it is also important to analyze the possible functional variable profile between distant related species (e.g. tetrapods vs teleosts) in the orthologous proteins by implementing the same strategy for ER alpha and beta paralogous proteins. The ER alpha-mediated reporter gene transactivation experiments have indicated that there is no major difference between mammal, bird and amphibian species[37]. But, the thermo-dependent ligand binding and gene expression differences are also observed for human and trout ER alpha proteins[12,37]. The computational analysis for detecting the functional divergence between the different clusters within the orthologous protein is very helpful to understand the evolution process in a protein family. Although the null hypothesis assumed the orthologous proteins have the same function specificity, and the paralogous proteins have the altered specificity[38]. We have not compared the heterogeneous substitution rates between orthologous proteins (tetrapods vs teleosts) with that of between paralogous proteins (alpha vs beta) in the case of ER. As Gribaldo *et al.* (2003)[35] suggested, the limited sampling will be biased to test the heteotachous sites in the protein family. We believe the enough sampling in the orthologous proteins within ER alpha or beta will improve the analysis of heterogeneous substitution rates between different clusters. Also, this analysis will be useful to discriminate the functional sites from general heterotachous sites.

References

1. Kuiper G.G., Shughrue P.J., Merchenthaler I., Gustafsson J.A. (1998) The estrogen receptor beta subtype: a novel mediator of estrogen action in neuroendocrine systems. Frontiers in Neuroendocrinology. 19: 253-286.
2. Pettersson K., Gustafsson J.A. (2001) Role of estrogen receptor beta in estrogen action. Annual Review of Physiology. 63: 165-92.
3. Nilsson S., Makela S., Treuter E., Tujague M., Thomsen J., Andersson G., Enmark E., Pettersson K., Warner M., Gustafsson J.A. (2001) Mechanisms of estrogen action. Physiological Reviews. 81 (4): 1535-1565.
4. Menuet A., Pellegrini E., Anglade I., Blaise O., Laudet V., Kah O., Pakdel F. (2002) Molecular characterization of three estrogen receptor forms in zebrafish: binding characteristics, transactivation properties, and tissue distributions. Biology of Reproduction. 66 (6): 1881-92.
5. Kelley S.T., Thackray V.G. (1999) Phylogenetic analyses reveal ancient duplication of estrogen receptor isoforms. Journal of Molecular Evolution. 49 (5): 609-14.
6. Hawkins M.B., Thornton J.W., Crews D., Skipper J.K., Dotte A., Thomas P. (2000) Identification of a third distinct estrogen receptor and reclassification of estrogen receptors in teleosts. Proceedings of the National Academy of Sciences of the United States of America. 97: 10751-10756.
7. Lassiter C.S., Kelley B., Linney E. (2002) Genomic structure and embryonic expression of estrogen receptor beta a (ERβa) in zebrafish (*Danio rerio*), Gene. 299 (1-2): 141-51.
8. Kuiper G.G., Enmark E., Pelto-Huikko M., Nilsson S., Gustafsson J.A., (1996) Cloning of a novel receptor expressed in rat prostate and ovary. Proceedings of the National Academy of Sciences of the United States of America. 93: 5925-5930.
9. Kuiper G.G, Carlsson B., Grandien K., Enmark E., Haggblad J., Nilsson S., Gustafsson J.A., (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta, Endocrinology, 138: 863-870.
10. Lindberg M.K., Moverare S., Skrtic S., Gao H., Dahlman-Wright K., Gustafsson J.A., Ohlsson C. (2003) Estrogen receptor (ER)-beta reduces ER alpha-regulated gene transcription, supporting a "Ying Yang" relationship between ER alpha and ER beta in mice. Molecular Endocrinology. 17 (2): 203-8.
11. Sun J., Baudry J., Katzenellenbogen J.A., Katzenellenbogen B.S. (2003) Molecular basis for the subtype discrimination of the estrogen receptor-beta-selective ligand, Diarylpropionitrile. Molecular Endocrinology. 17 (2): 247-58.
12. Matthews J.B., Clemons J.H., Zacharewski T.R. (2001) Reciprocal mutagenesis between human alpha (L 349, M528) and rainbow trout (M317, I496) estrogen receptor residues demonstrates their importance in ligand binding and gene expression at different temperatures. Molecular and Cellular Endocrinology. 183: 127-139.
13. Knudsen B., Miyamoto M.M., Laipis P.J. and Silverman D. N. (2003) Using evolutionary rates to investigate protein functional divergence and conservation: A case study of the carbonic anhydrases. Genetics. 164 (4): 1261-9.
14. Lesk A.M. (2001) Introduction to Protein Architecture Oxford University Press, 171-175.
15. Gu X. (1999) Statistical methods for testing functional divergence after gene duplication. Molecular Biology and Evolution. 16: 1664-1674.
16. Gu X. (2001) Maximum-likelihood approach for gene family evolution under functional divergence. Molecular Biology and Evolution. 18: 453-464.
17. Knudsen B., and Miyamoto M.M. (2001) A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. Proceedings of the National Academy of Sciences of the United States of America. 98: 14512-14517.
18. Blouin, C., Boucher Y., and Roger A.J. (2003) Inferring functional constraints and divergence in protein

families using 3D mapping of phylogenetic information. Nucleic Acids Research. 31: 790-797.
19. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25 (17): 3389-402.
20. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 25 (24): 4876-82.
21. Kumar S., Tamura K., Jakobsen I.B., Nei M. (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics, 17 (12): 1244-5.
22. Strimmer K., and Von Haeseler A. (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Molecular Biology and Evolution. 13: 964-969.
23. Gu X., Vander Velden K. (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics. 18 (3): 500-1.
24. Humphrey, W.F., Dalke A., and Schulten K. (1996) VMD-Visual Molecular Dynamics. Journal of Molecular Graphics. 14: 33-38.
25. Drouin G. (2002) Characterization of the gene conversions between the multigene family members of the yeast genome. Journal of Molecular Evolution. 55 (1): 14-23.
26. Takahashi K., and Nei M. (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Molecular Biology and Evolution. 17 (8): 1251-8.
27. Shiau A.K., Barstad D., Loria P.M., Cheng L., Kushner P.J., Agard D.A., Greene G.L. (1998) The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen, Cell. 95 (7): 927-37.
28. Pike A.C., Brzozowski A.M., Hubbard R.E., Bonn T., Thorsell A.G., Engstrom O., Ljunggren J., Gustafsson J.A., Carlquist M. (1999) Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. The EMBO Journal. 18 (17): 4608-18.
29. Eng F.C., Lee H.S., Ferrara. J., Willson T.M., White J.H. (1997) Probing the structure and function of the estrogen receptor ligand binding domain by analysis of mutants with altered transactivation characteristics. Molecular and Cell Biology. 17 (8): 4644-53.
30. Humeny A., Bokenkamp D., Thole H.H. (1999) The HDQVH-motif in domain E of the estradiol receptor alpha is responsible for zinc-binding and zinc-induced hormone release. Molecular and Cellular Endocrinology. 153 (1-2): 71-8.
31. Landgraf R., Xenarios I., Eisenberg D., (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. Journal of Molecular Biology. 307 (5): 1487-502.
32. Maeda M. (2001) The conserved residues of the ligand-binding domains of steroid receptors are located in the core of the molecules. Journal of Molecular Graphics and modeling. 19 (6): 543-51.
33. Lichtarge O., Bourne H.R., Cohen F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families. Journal of Molecular Biology. 257 (2): 342-58.
34. Philippe H., Lopez P. (2001) On the conservation of protein sequences in evolution. Trends in Biochemical Sciences. 26 (7): 414-6.
35. Gribaldo S., Casane D., Lopez P., and Philippe H. (2003) Functional Divergence Prediction from Evolutionary Analysis: A Case Study of Vertebrate Hemoglobin. Molecular Biology and Evolution. 20 (11): 1754 -9.
36. Lichtarge O., Yamamoto K.R., Cohen F.E., (1997) Identification of functional surfaces of the zinc binding domains of intracellular receptors. Journal of Molecular Biology. 274 (3): 325-37.
37. Sumida K., Ooe N., Saito K., Kaneko H. (2003) Limited species differences in estrogen receptor alpha-medicated reporter gene transactivation by xenoestrogens. The Journal of Steroid Biochemistry and Molecular Biology 84 (1): 33-40.
38. Li L., Shakhnovich E.I., Mirny L.A. (2003) Amino acids determining enzyme-substrate specificity in pro-

karyotic and eukaryotic protein kinases. Proceedings of the National Academy of Sciences of the United States of America 100 (8): 4463-8.

サイト別レイトシフト分析と3次元構造マッピングを用いた機能的重要残基の予測

王　峰[1]・岸野　洋久[1,2]・和田　康彦
(動物生産学分野)

平成16年10月27日　受理

エストロゲン受容体αとβ間のリガンド結合領域における機能の相違をサイト別レイトシフト法で分析した．エストロゲン受容体のサブタイプの機能変化に関わる機能的残基を事後確率と尤度比検定を用いて予測した．さらに，3次元構造マッピングを組み合わせることにより，リガンド結合であるか，2量体化のためのインターフェースであるかといった機能を，予測された残基について判別することが可能となった．この複合アプローチは実験におけるターゲットとなりうる機能的残基をスクリーニングすることが可能である．また，このアプローチによってエストロゲン受容体ファミリーのアミノ酸配列と機能の間の関連性についての新しい知見をもたらした．

[1]BIRD，科学技術振興機構
[2]東京大学大学院農学生命科学科生物測定学研究室